DOCUMENT RESUME

ED 247 234 TM 840 137

AUTHOR Jones, Douglas H.; And Others
TITLE A Statistical Method for Identifying Racial Overrepresentation in Special Education Placement.
INSTITUTION Educational Testing Service, Princeton, NJ. Program Statistics Research Project.
REPORT NO ETS-PSR-TR-83-42; ETS-RR-83-39
PUB DATE Aug 83
NOTE 46p.
AVAILABLE FROM Educational Testing Service, Publications Order Services, Dept. I-101, Princeton, NJ 08541.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Comparative Analysis; Elementary Secondary Education; Mental Retardation; *Minority Group Children; *Racial Discrimination; Sampling; *Special Education; State School District Relationship; *Statistical Analysis; Statistical Studies; *Student Placement

ABSTRACT

The discovery of disproportionate minority enrollments in special education classes has been a subject of considerable concern to federal, state, and local education officials alike for the past decade. Because the existence of disproportions is frequently thought to result from discriminatory placement practices on the part of local school districts, many state agencies are implementing methods for detecting and taking action against offending districts. This paper discusses some of the methodological problems inherent in existing methods for determining disparities in special education placement and offers an alternative to presently-used methods. The alternative offered can be used for the simultaneous comparison of disparity across many districts and several diagnostic categories. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED247234

TM 840 137

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

H. Weidenmiller

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

A Statistical Method for Identifying Racial Overrepresentation In Special Education Placement

Douglas H. Jones
Randy Elliot Bennett

and

Jerome Sacks
Northwestern University

Program Statistics Research
Technical Report No. 83-42

Research Report RR-83-39

August 1983

Not to be cited or further reproduced
without the permission of the authors.

Copyright © 1983 by Educational Testing Service. All rights reserved.

The Program Statistics Research Technical Report Series is designed to make the working papers of the Research Statistics Group at Educational Testing Service generally available. The series consists of reports by the statisticians in the Research Statistics Group as well as their external and visiting statistical consultants.

Reproduction of any portion of a Program Statistics Research Technical Report requires the written consent of the author(s).

Abstract

The discovery of disproportionate minority enrollments in special education classes has been a subject of considerable concern to federal, state, and local education officials alike for the past decade. Because the existence of disproportions is frequently thought to result from discriminatory placement practices on the part of local school districts, many state agencies are implementing methods for detecting and taking action against offending districts. This paper discusses some of the methodological problems inherent in existing methods for determining disparities in special education placement and offers an alternative to presently-used methods. The alternative offered can be used for the simultaneous comparison of disparity across many districts and several diagnostic categories.

Table of Contents

	Page
Introduction	1
Statistical Background and Critique	3
Random Sampling	10
Relevance of Statistical Tests	12
The 80% Rule	14
Separate District Comparison or Simultaneous Comparison?	17
Simultaneous Comparison (1)	19
Appendix	28
Conclusion	29
References	30

The possibility of cultural and racial discrimination in the special education evaluation and placement process has become a matter of serious concern to courts, federal and state agencies, and local school districts. The observed fact that, relative to their proportion in the general school-age population, minority children are overrepresented in special education, particularly in classes for the "educably mentally retarded" (EMR), has led to extensive debate about the causes for this disparity (Prasse, 1980; Reschly, 1979). Evidence of this debate is found in the monitoring by government agencies of the rates of placement, mandates for culture-fair assessment in special education legislation and regulatory codes (U.S. Office of Education, 1977; Department of Health, Education, and Welfare, 1977), the establishment of a National Academy of Sciences Panel (Heller, Holtzman, and Messick, 1982) to study the problem and civil suits alleging discriminatory practices (Larry P. v. Wilson Riles, 1979; Isaac Lora v. Board of Education, City of New York, 1979; PASE v. Hannon, 1980; Mattie T. v. Holladay, 1979).

Evidence for determining discriminatory behavior in the placement of EMR children begins with the implicit presumption that true mental retardation is equally distributed across ethnic groups. This presumption can be argued since many of the factors thought to contribute to mental retardation (e.g., malnutrition, poor maternal and developmental health care, limited educational opportunity) disproportionately affect minority groups. All the same, in litigation, evidence of a disproportionate number of minority children assigned to EMR classes is often sufficient to establish a prima facie case and require the defendants to validate their placement procedures. Some state agencies have formulated regulations about the assessment of such data and

school districts found in violation are then required to investigate and remedy possible causes of the disparity. Thus, a finding, based solely on data, that there is a disparity has considerable consequences.

The use of statistical methods for assessing the data has been recognized by governmental agencies and courts; in analogous situations involving employment discrimination, the U.S. Supreme Court has accepted the use of statistical procedures though it has not made clear how gross a statistical disparity must be to establish a prima facie case (Hazelwood School District v. United States, 1977).

The use of standard statistical methods on placement data is predicated on assumptions which are almost certain to be violated, thereby calling into question whether such methods are relevant for assessing disparity. Unthinking application of statistical significance tests can lead to detection of even minute differences if the individuals in the samples are numerous enough. Agencies which monitor hundreds of school districts and automatically apply a simple statistical procedure to the data from each district may find, with large probability, a sizeable number of school districts whose procedures are culture fair and nondiscriminatory, but which appear to have a "statistically significant" disproportionate number of minority children in EMR classes. Then there is an inherent conflict between assuring fair treatment of minorities and unduly burdening innocent school districts. Moreover, in order to comply with the demands of such analyses, school districts will inevitably be forced into quota systems of assignment with consequences that may be at variance with the purpose of special education programs.

Our goal in this article is to explain the statistical issues raised, to criticize some of the proposed statistical methods in use, and to offer an approach which can serve to detect disparities, and, simultaneously, avoid the pitfalls of methods presently used. While there is much to be concerned about the role of statistical methodology in assessing special education placement, our view is that it can serve as a useful benchmark when exercised with care. Our approach is determined by the need to simultaneously assess the data from a large number of districts. Such issues have been treated in the statistical literature and what we propose is not a new statistical methodology but an adaptation of existing methods to the context of special education placement.

Statistical Background and Critique.

In this section, we describe the statistical analyses commonly used to assess disparity in placement rates, the assumptions on which the methodology is based and a discussion of the appropriateness of such methodology.

Significance Tests.

In Larry P v. Riles data was reported comparing proportions of black enrollment in EMR classes to proportions of non-black enrollments. Thus, in Oakland there were (in 1976-1977) 633 blacks enrolled in EMR classes and 88

- - - - - - - - -

Table 1 about here

- - - - - - - - -

non-blacks (we shall refer to them as whites); the school population consisted of 35,086 blacks and 17,189 others. The rate of black EMR was 663/35,086 = .0189 and the rate of white EMR was 88/17,189 = .0051. The opinion notes that

TABLE 1

1976-1977 ENROLLMENT DATA ON TWENTY SCHOOL DISTRICTS WHICH ACCOUNT FOR 80% OF CALIFORNIA'S BLACK POPULATION*

District	EMR Enrollmt.	Black EMR	Total District Enrollmt.	Total Black Enrollmt.	Black District %	Black EMR %	Non-Black EMR Enrollmt. Rate	Black EMR % Enrollmt. Rate
L.A.U.S.D.	3155	1786	592,931	142,778	24.1%	56.6%	.30%	1.25%
Oakland	751	663	52,275	35,086	67.1%	88.3%	.51%	1.89%
Compton	461	442	31,054	26,356	84.9%	95.9%	.40%	1.68%
S.F.U.S.D.	446	240	69,982	21,730	31.1%	53.8%	.43%	1.10%
San Diego	592	202	121,233	17,637	14.6%	34.1%	.38%	1.14%
Richmond	109	67	34,880	12,278	35.2%	61.5%	.19%	.54%
Pasadena	38	29	25,409	10,872	42.8%	76.3%	.06%	.27%
Inglewood	62	51	13,068	10,428	79.8%	82.3%	.42%	.49%
Sacramento City	NO DATA							
Long Beach	458	75	60,187	9,058	15.1%	16.4%	.75%	.83%
Pomona	139	77	20,863	6,195	29.7%	55.4%	.42%	1.24%
Fresno City	399	110	53,643	5,550	10.4%	27.6%	.60%	1.98%
Berkeley	25	22	12,142	5,221	43.0%	88%	.04%	.42%
San Bernardino City	265	79	30,970	4,712	15.2%	29.8%	.02%	1.68%
Lynwood	94	56	9,450	4,560	48.3%	59.6%	.78%	1.23%
Stockton City	260	84	26,868	4,293	16.0%	32.3%	.78%	1.96%
Vallejo	81	39	14,357	4,030	28.1%	48.2%	.41%	.97%
Ravenswood	10	9	3,443	2,913	84.6%	90%	.19%	.31%
Bakersfield	185	73	18,346	2,893	15.8%	39.5%	.72%	2.52%
Monterey Peninsula	90	36	14,003	2,627	18.8%	40%	.47%	1.37%

*SOURCE: Larry P. Table 2, page 306.

the difference in rates is statistically significant at the five percent level.

The basis for the last statement is a statistical comparison of two rates -- the two-sample binomial test where the two samples are the black and white children of the district. The comparison of the two rates is done by first computing the overall EMR enrollment rate equal to 751/52,275 = .0144, then computing the standard error, SE, of the difference of the rates,

$$SE = ((.0144)(.9856)\left(\frac{1}{35086} + \frac{1}{17189}\right))^{1/2} = .0011,$$

and, finally, using the fact that, if there was no difference in the "true" enrollment rates of the two groups, the random fluctations of the difference in rates/SE is governed (at least to a high degree of approximation) by the normal distribution. Thus, we compute

$$Z = \text{the observed difference in rates/SE}$$
$$= (.0189 - .0051)/.0011 = 12.44,$$

and note that the probability of obtaining a fluctuation resulting in such a large value of Z when, in fact, there is no difference in the "true" enrollment rates is, from tables of the normal distribution, virtually 0. This probability is called the significance probability (or P-value); it being less than .05 in the instant case accounts for the court's statement that the difference is significant at the five percent level.

For the Inglewood school district we note that the placement rates are .0049 and .0042 for blacks and whites respectively. The overall rate is 62/13068 = .0047 and the standard error is

$$SE = ((.0047)(.9953)(\frac{1}{10428} + \frac{1}{2640}))^{1/2} = .0015$$

and

$$Z = (.0049 - .0042)/.0015 = .4692$$

giving a significance probability, from tables of the normal distribution, of .64, much greater than the five percent standard. As noted by the court the difference in rates in Inglewood is not significant.

An alternate way of proceeding would be to regard the EMR students as a random sample from an undetermined population of EMR eligibles. Assuming that the population of EMR eligibles is racially constituted in the same way as the entire school population so that in Oakland the proportion of black EMR "eligibles" is 67.1% we could ask whether the proportion of blacks in the observed EMR group (for Oakland it is 88.3%) is significantly different than the expected 67.1%. Such an analysis is, in fact, conceptually simpler than the earlier two-sample method because we are now, in effect, dealing with a single sample. Such a one-sample analysis proceeds by calculating the appropriate standard error

$$SE = ((.671)(.329)/751)^{1/2} = .017$$

and

$$Z = (.883 - .671)/.017 = 12.37$$

numerically only slightly different than the Z-value calculated for the two-sample analysis earlier and leading to the same qualitative result: a significance probability indistinguishable from 0. A similar calculation for Inglewood produces an SE = .051 and Z = (.823 - .798)/.051 = .49.

Although both presented models lead to essentially equivalent conclusions we must inquire which is more appropriate, not to make overly fine distinctions, but to clarify interpretations of results and to distinguish between alternative analyses where different conclusions may differ dramatically depending on the model. One such alternative, used by the State of Massachusetts to monitor school districts, finds disparity if the white enrollment rate is less than 80% of the black enrollment rate. Applied to representation rates such a rule can lead to entirely different conclusions. For example, in Table 1 this 80% rule finds the Ravenswood enrollment rates disparate (.0009 is less than .8 x .0031 = .0024) but finds the opposite when applied to representation rates (.846 is greater than .8 x .9 = .72). A similar discrepancy occurs in the Compton data. We shall say more about this 80% rule later.

Which model to use brings forward the issue of whether the enrollment rates or the representation rates better reflect the impact of the placement process. Enrollment rates have as reference a universe of school children wide enough to admit belief that the district school children are a random sample from that universe; representation rates refer to a more narrow universe of EMR "eligibles" within the school district.[1]

---

[1]The reason both models lead to such similar results as they do in the analysis of the Oakland data is worth noting: let E_B, E_W be the number of blacks and whites enrolled in EMR and let N_B and N_W be the number of blacks and whites in the school district. The enrollment rates are E_B/N_B and E_W/N_W for blacks and whites respectively while the representation rates are $E_B/(E_B+E_W)$ and $N_B/(N_B+N_W)$ for EMR and district respectively. If the enrollment rates are equal ($E_B/N_B = E_W/N_W$) then it must be that the representation rates are also equal and vice versa. Thus differences in one set of rates reflects the differences in the other set of rates — the differences go in the same direction but to a different degree.

Use of enrollment rates proceeds by posing the principle: if the incidence of EMR in the universe is the same for blacks and whites then, the school district being a random sample, the incidence (enrollment) rates in the district should reflect this. If the district rates differ substantially then either the assumption of random sampling is suspect or the assumption of equivalence of EMR incidence for the two races is untenable. Belief in equivalence of the rates in the universe then puts the issue of random sampling to the front. Then the connection to be made is between the placement process and the assumption of random sampling.

A difference in district incidence rates can be recognized as a result of the placement process provided each of the individuals in the district are "examined" for EMR. Thus, all those who are <u>not</u> placed in EMR are assumed to have been diagnosed "non-EMR". However, actual placement practices have no such characteristic. Children who are placed in EMR are first referred and then evaluated for placement; that all children not referred are thereby diagnosed as non-EMR is not a credible supposition. Referrals occur because of disruptive behavior, failure to perform, etc.; others, for whatever reason, do not "select" themselves for attention and never enter the process. Moreover, evaluation is made after referral and cannot be investigated without consideration of the referrals. Thus, the model using enrollment rates is a dubious one for assessing effects of the placement process, not only does it fail to distinguish the referral and evaluation stages but it supposes that the individuals in the district form a random sample and that each individual is diagnosed for EMR.

The placement process in a district must then be studied by examining

stages at which children are "processed". Referrals for evaluation are made from some subset of the district's children, a subset we call the EMR "eligibles". The characteristics of the set of eligibles cannot, with present knowledge, be adequately described. We are then faced with a hierarchy of sets of children:

1) district
2) EMR eligibles
3) referrals
4) EMR.

This is analogous to employment situations where the hierarchy is 1) working population, 2) eligible work force, 3) applicants, and 4) hires.

With this hierarchy in mind we regard the placement process in two stages: the referral stage and the evaluation stage. Analysis of the evaluation stage could proceed, assuming data for referrals was available, by comparing the proportions of EMR blacks among black referrals to the similarly calculated proportion for whites. Such an analysis in employment situations compares the rate of hire among black applicants to the rate for whites and is the common method employed when "applicant flow" data is available and pertinent. When referral data is unavailable or misleading, as it might be because of varying (perhaps discriminatory) practices, accurate assessment of the evaluation stage is out of the question.

If referral data is available, assessment of the referral stage requires knowledge of the composition of the eligible set. A common assumption is that the needed composition is the same as that of the district and then comparisons can be made between the proportion of blacks among the referred

and the proportion in the district. Leaving aside questions about the appropriateness of the assumption, the results from analyses of the two stages serve to narrow attention to possible sources of disparity. Thus, if the referral rate for blacks is .8333 (such numbers are reported in Ashhurt and Meyers (1973)) then disparity at the evaluation stage is not indicated but there appears to be a disproportionate number of blacks referred. The question then centers on whether such results, by themselves, should lead to a prohibition of tests (such as IQ) used at the evaluation stage to place students in EMR.

It is important to note that appropriateness of IQ tests is not directly in question — the only issue is whether the use of such tests has an adverse impact on blacks. In light of Connecticut v. Teal (1982), it is uncertain whether the "bottom-line" approach, which calls into question any aspect of the entire placement process if overall disparity is present can be defended.[2] A discussion of IO tests in the placement process has been given by Lambert (1981) who reports data indicating that elimination of IQ tests does not produce different overall placement proportions, perhaps indicative of the controlling effect of the referral stage.

When referral data is unusable or unavailable a possible analysis of placement rates is to compare the representation of blacks in EMR with their representation in the eligible set. Assuming that the latter has the same racial composition as the district we reach the companion of representation

---

[2] In Teal a test for promotion was prohibited because it had an adverse impact on blacks despite the fact that blacks were eventually promoted in greater numbers, proportionately, than whites. Whether the opposite situation, as in our example, would be treated symmetrically is not clear.

rates described earlier for the Oakland data. Support for this type of analysis for establishing a prima facie case may be found in International Brotherhood of Teamsters v. U.S. where the Supreme Court states that absent explanation, it is ordinarily to be expected that nondiscriminatory hiring practices will in time result in a work force more or less representative of the racial and ethnic composition of the population in the community from which employees are hired.

The defensibility of this position will be examined in the next section. The immediate concern about which rates to compare, enrollment or representation, is resolved in favor of representation rates because the model for their comparison reflects more closely the placement process and focuses more directly on sources of observed disparities.

Random Sampling

Use of statistical significance tests requires assumptions of random sampling — in the first analysis described above the black and white school populations are two random samples; in the second analysis the EMR students are a random sample from the population of EMR "eligibles". It is not immediately clear that such an assumption is tenable. Assuming random sampling in any statistical context is difficult, but a placement process, inherently subject to many influences, can never result in a random sample. We discuss two plausible phenomena that violate the assumptions of random sampling.

Strata

First, suppose placement decisions are made separately at two groups of schools in a district where one group of schools has 8,000 blacks and 800 whites of which 160 blacks and 16 whites are placed in EMR, while the other

group of schools has 4,000 blacks and 4,000 whites of which 40 of each race are EMR. The EMR placement rate in each group shows no impact on blacks (in the first group, two percent of each race are EMR, in the second group one percent). Aggregation of the data induces an appearance of disparity since 200 of the 12,000 blacks are EMR (a placement rate of .017) while 56 of the 4,800 whites are EMR (a placement rate of .0117). Indeed, the two-sample test produces a significance probability of .01, well below the five percent standard. If the data were analyzed as a one-sample comparison of the population of blacks in EMR (200/256 = .7813) compared to the proportion of blacks in the district school population (12,000/16,800 = .7143) then the result would be a significance probability of .018.

The phenomenon that aggregating data can lead to results contrary to that found for disaggregated data is sometimes known as Sampson's Paradox (Bickel, Hammel, and O'Connell, 1975). In our example, the "paradox" stems from the existence of two strata and the consequent failure of the random sampling hypothesis.

When strata are easily recognized, as they might be in the example, they can be taken into account. However, stratification may occur in non-detectable ways to an unknown extent and defy incorporation into a model for analysis. Even when strata are detected, disaggregation can be misleading (c.f. Kruskal, 1977). In short, it is implausible that strata effects can be compensated for and their presence determines the random sampling assumption necessary for application of the statistical significance tests.

Clusters

Second, if the selection of individuals into the sample occurs in clusters; then, the measured attributes between individuals are statistically

dependent. For example, siblings typically go to school in the same school district or, the placement of one child in EMR may result in a higher chance for another child being so placed if the second child is related to the first then if there were no relationship. What results is a lack of statistical independence caused by the absence of random sampling and doubt is then cast upon the validity of the statistical tests.

Any number of other factors can cause perturbation in the random sampling assumption. Placement practices may vary over time, the demographic composition of the district may vary over time, etc. In short, random selection in the special education placement process is not an acceptable hypothesis.

Relevance of Statistical Tests

In light of the difficulties cited under random sampling it is potentially misleading to rely solely on statistical significance tests to determine the adverse impact on minorities of a placement process. There are many who would ignore such caveats and proceed with the statistical analysis (perhaps due to belief that any quantification is better than none) and there are voices who would reject the use of statistical analysis in such circumstances ("the calculations are meaningless if the assumptions are not met"). Our own position is less extreme: the use of statistical analysis can be a valuable benchmark. If, under the assumption of a benchmark random sampling process, there is no statistically significant difference, then we would not believe that the data provides evidence of discrimination. If, on the other hand, statistical significance is found, it is premature to state that disparity exists for two reasons: first, the random sampling model is

suspect; second, the disparity may not be substantial or of "practical" significance. We should be led, therefore, to make some allowance for departures from the random sampling model and incorporate some measure of practical significance. At the same time we should be careful to rely on such an approach in a context free of gross departures from the benchmark random sampling model.

Consideration of statistical evidence for determining disparity in special education placement must begin with the question of what ought to be expected when the placement process is nondiscriminatory. The common belief that nondiscriminatory behavior should lead to equal placement rates for blacks and whites or, the proportion of blacks in EMR should be equal to the proportion of blacks in the school population is indefensible in light of the fact that groups defined by any characteristic (e.g., EMR) are almost certain to differ on other characteristics (e.g., race) without presupposing the direction or the magnitude of the difference. An unpleasant feature of statistical tests is that, with large enough sample sizes, even a small and unimportant difference will be found statistically significant. For example, suppose that the number of black EMR's in Oakland is 228 rather than 663. Then the two-sample tests described earlier would find significance at the five percent level while the the black EMR enrollment rate is .65% compared to .51% for others, and the proportion of blacks in EMR classes is 72.2% compared to 67.1% in the school population. Whether such differences in percentages can be regarded as practically significant is not self-evident nor is it clear that considering the almost certain presence of differences in groups such disparities can properly be regarded as evidence for discriminatory behavior.

The 80% Rule

The State of Massachusetts Department of Education monitors school districts and those in which the white enrollment rate in special education is less than 80% of the black or hispanic enrollment rate are cited for prima facie derial of equal educational opportunity. This 80% rule is similar to, but not the same as an 80% rule initiated by the EEOC and incorporated in the Uniform Guidelines for Employed Selection Procedures (1978) which has been adopted by federal agencies. The agency guidelines find that an employee qualification test has an adverse impact on, for example, blacks, if the black pass rate (number of blacks who pass/number of blacks who take the test) is less than 80% of the white pass rate and if the disparity is statistically significant (usually by use of a two-sample test for the difference of the two pass rates). We note that the Massachusetts version of the 80% rule does not incorporate the significance test.

Meier, Sacks, and Zabell (1982) in their analysis of the EEOC 80% rule, have noted that it is preferable to the sole use of a significance test or to the sole use of the 80% component. The two components of the 80% rule serve as an assessment of whether the results of the selection procedure could have reasonably arisen by chance (the significance test) and, via the 80% component, offers some allowance for departure from the random sampling assumption as well as a measure of practical significance. The Massachusetts procedure lacks an assessment of the random aspects of placement thereby placing such narrow restrictions that school districts, in order to avoid citation, must employ a placement process virtually equivalent to a quota system. For example, in a school district with 6,000 of each race, and EMR

enrollment rates of .01 and .0075 for blacks and whites respectively the z-score computed is (.01 -.0075), /√(.00875 x .99125 x 2/6000) = 1.47 which gives a significance probability of .14. If the numbers of blacks and whites are 10,660, respectively, then the significance probability would be .05. Generally, unless the numbers of blacks and whites are very large, statistical significance will be absent when the ratio of white EMR enrollment rate to black EMR enrollment rate is close to .80. Of course, disparities considerably greater than the 80% standard will typically be associated with statistical significance (as in most of the districts in Table 1). Our point here is that adherence to an 80% rule of the type utilized in Massachusetts can result in misleading assessments of the random elements.

The chance is substantial that a fair process will be found disparate by sole application of the 80% component. For example, in a district with 5,000 of each race and an overall EMR placement of .005 the probability is about .20 that an 80% disparity would occur; if the EMR placement rate is .008 then the probability is about .16. Thus, one can expect a proportion of the 16% to 20% of non-discriminatory moderate-sized districts to be cited by the Massachusetts plan. Whether this burden on school districts is acceptable or desirable is a question we do not address.

We have argued earlier that the use of enrollment rates is not appropriate; the relevant population should be EMR "eligibles" and the comparison, lacking further information, could be done on black representation rates in EMR classes and districts. An 80% rule can be utilized in this context to help remedy some of the ailments resulting from the sole use of

statistical significance tests. Such a rule would find disparity if the black district rate is less than 80% of the black EMR rate (not the enrollment rate) and the disparity is statistically significant. There is a profound difference in the way the 80% component behaves here for two reasons. First, the sample sizes and rates are very different; the number of EMR students now determine the sample size rather than the number of students in the district, resulting in different standard error calculations. This has an effect on the statistical significance test and accounts for the somewhat different z-scores calculated earlier. Second, the rates on which the 80% component are calculated are different; the enrollment rates are very small numbers but the district and EMR rates range, in Table 1, between ten percent and 95% and, especially at the higher representation rates, the effect of the 80% calculation can be noticeable. Applied to the data of Table 1, the districts which meet the 80% standard are Compton (.80 x .959 = .767 which is less than .849), Inglewood, Long Beach, Lynwood, and Ravenswood. Of these the Inglewood, Long Beach, and Ravenswood disparities are not statistically significant while Compton and Lynwood are. Note also that the sole application 80% component to the enrollment rates for Ravenswood would find disparity. Lynwood is marginal in that .80 x .596 = .477 which is less than .483 but close to it. Thus, strict application of both components of the 80% rule would find disparity in all but five districts in Table 1.

In summary: (1) Sole use of statistical significance tests is not defensible because the necessary assumption of random sampling is unwarranted and because it lacks a measure of practical significance. (2) Use of an 80% standard alone can be misleading because it does not account for random

fluctuations. (3) Comparison of enrollment rates is inappropriate because it does not deal with the relevant propulations. (4) A simple rule which responds to some of these issues is to take as the relevant sample, the EMP students and to apply the EEOC 80% rule (both the 80% standard and the significance test) to the black EMR rate.

Separate District Comparison or Simultaneous Comparison?

The question addressed by the court in Larry P. is whether the use of intelligence tests for EMR placement has adverse impact on black school children in the State of California i.e., on all school districts in the state. The question arises as to whether assessing the data from each school district separately is a proper means of assessing state-wide practices.

Consider a hypothetical example where 20 districts, all acting fairly, are monitored and a district is flagged if the difference between the black EMR rate and the black district rate is positive and more than 1.96 standard error. The probability that none of the 20 districts are flagged is $(.975)^{20} = .603$ or, with probability about .40, some districts will be flagged. In fact the probability is .309 that exactly one will be flagged, .075 that exactly two will be flagged, .01 that exactly three will be flagged, etc. The reason, of course, is that if many significance tests are run, each having .025 probability of finding a disparity unfavorable to blacks, then some of these tests, purely by chance, will turn out significant. Should this process be applied to 200 or 2,000 districts then one can expect the citation of substantial numbers of innocent districts. Thus, assessing district by district, while appearing reasonable for each district, has serious effect on

an overall assessment.

In Larry P. one of the court imposed remedies is that any district in California in which the disparity in enrollment rates was greater than one standard deviation had to submit a plan for correcting the balance. Thus, any placement process which, on long-run usage, would place blacks and whites in EMR proportionately, has a 16% chance of being challenged (this is because the probability of a z-score greater than 1 is .16) unless conscious effort was made by the district to adjust placements in order to meet the one standard deviation benchmark. In effect, this pressures school districts to adopt a quota system of placement. A similar effect ensues under the Massachusetts procedure or, in fact, if any of the single district procedures discussed in the previous section are employed to monitor districts. Whether special education purposes or the court's intent are then served are matters we do not discuss.

The comparisons made in Larry P. were used as evidence that EMR placement practices impacted adversely on black children. The evidence seems strong — fully three-fourths of the districts studied show large disparities according to the application of the EEOC 80% rule described earlier. There is no guidance from the courts about how strong the statistical evidence should be to establish a prima facie case of discrimination and, in settings like that of Larry P. it is not clear whether evidence showing two out of 20 districts disparate, or five out of 20, is enough. Which index to use cannot be supplied by mathematics, though the statistical properties of a specific index might illuminate matters. For example, if the index is two out of 20 disparate then according to the calculations made earlier in this section,

there is an 8.8% chance of finding at least two disparate districts out of 20 if single district comparisons are made even if all the districts' placement procedures were nondiscriminatory and the data could be regarded as coming from random samples.

Thoughtless combinations of single comparisons thus have inherent flaws and it is necessary to employ methodology which directly addresses the multiplicity of comparisons.

Simultaneous Comparison (1)

The need to use a methodology appropriate for simultaneous analysis of all twenty districts in Table 1 leads us to a procedure known to statisticians but, perhaps, not widely enough known in educational applications (c.f., Miller, 1981). We compute all 20 z-scores for comparing the black EMR rate with the black district rate. The significant z-scores will be those which exceed a critical value, CRIT, to be determined. The value CRIT should not be 1.96 lest we wish to ignore flagging innocent districts with substantial probability. If CRIT = 3.02 then, assuming that the largest of the 20 z-scores would exceed CRIT purely by chance, the probability is .025 (we choose half the five percent significance level because we are only interested, at this point, in detecting disparities adverse to blacks). The details of this simple calculation are in the Appendix.

Applying this to the data of Table 1 we note that Inglewood, Long Beach, Lynwood, and Ravenswood have z-scores less than 3.02 (Lynwood's is 2.19). Of the ones remaining application of the 80% rule rules out Compton. It is worth noting the effect of the simultaneous comparison: Lynwood which only

marginally satisfied the 80% rule and had a statistically significant disparity based on the one-sample test would not be found disparate in the simultaneous analysis; the three districts which did not exhibit significant differences would remain insignificant.

If the number of districts to be monitored is very large then CRIT becomes large because we are avoiding declaring even one innocent district guilty. It is then possible that some guilty districts would go undetected. For example, if we had 740 districts and CRIT=4 then the probability of getting at least one z-score greater than 4, when there are no differences, is .025. Then, districts with 150 EMR students, nine black EMR rates of .50, and a black district rate of .35 would go undetected since their z-scores = $.15/(.35 \times 65/150)^{1/2} = .385$. It is impossible to have the number of districts large, CRIT reasonably small (say 3 or less), and the probability of citing an innocent district at the .025 level. We are then faced with some trade-offs. If we increase the .025 level we can keep CRIT from getting large. We must then be prepared to state how many falsely cited innocent districts are tolerable.

Table 2 lists, for some values of CRIT and some numbers of districts in the analysis, the number of innocent districts (x) and the probability (P) that at least x would be cited if all districts were acting fairly and the incidence of EMR is the same for blacks and whites.

- - - - - - - - - - -

Table 2 about here

- - - - - - - - - - -

TABLE 2

PROBABILITY OF DETECTING AT LEAST x DISTRICTS FROM A TOTAL OF y HOMOGENEOUS DISTRICTS

CRIT=2.58

x	y→ 20	100	250	500	1000
1	0.0954	0.3942	0.7144	0.9148	0.9933
2	0.0045	0.0898	0.3556	0.7135	0.9599
3	0.0001	0.0141	0.1311	0.4565	0.8760
4		0.0017	0.0379	0.2422	0.7357
5		0.0002	0.0089	0.1083	0.5599
6			0.0018	0.0416	0.3840
7			0.0003	0.0139	0.2374
8				0.0041	0.1328
9				0.0011	0.0676
10				0.0003	0.0315

CRIT=3.00

x	y→ 20	100	250	500	1000
1	0.0257	0.1220	0.2776	0.4782	0.7277
2	0.0003	0.0077	0.0425	0.1385	0.3732
3		0.0003	0.0045	0.0282	0.1428
4			0.0004	0.0044	0.0430
5				0.0006	0.0106
6				0.0001	0.0022
7					0.0004
8					0.0001

How the results of such an analysis ought to be interpreted by the courts is not self-evident. For the data of Larry P. 15 of the 20 districts are disparate indicating wide-spread disparity. Had only three districts been found disparate it is not clear if the evidence would support a finding of state-wide disparity. The answer cannot be found in mathematics or principles of statistical reasoning. However, a specific benefit from the simultaneous analysis is the pinpointing of those districts which have disparities thereby permitting judicial and administrative focus on the identified problem areas.

Simultaneous Comparisons: An Example Using a Single District

In the previous section we have given a method for simultaneously analyzing the impact of EMR placement on blacks in a number of districts. Problems connected with special education placement often involve more than two groups (in many districts there are three major groups: blacks, hispanics, and whites) and many placement categories (Emotionally Distrubed (ED), Perceptually Impaired (PI), etc.). It is then important, both for judicial and administrative purposes, to have a methodology which can single out the groups and categories where disparities exist in order to uncover possible discriminatory behavior. The methods used in Section 4 can be adopted to these more complex situations and while the details are likewise more complex, the principle is the same.

To make matters specific suppose we have three groups and several categories and, for the moment, we restrict attention to a single district for which we present some actual data (see Table 3 for data and Table 4 for category codes).

- - - - - - - - - - - - - -

Tables 3 and 4 about here

- - - - - - - - - - - - - -

In this district the proportions of whites, blacks, and hispanics are 0.2169, 0.4989, and 0.2850 respectively. We calculate a z-score for each of the 21 cells in Table 3 as follows. For each category and group compute the number of group members in category/category total = $\hat{p}_{cg}$ (for example, $\hat{p}_{EMR,\ BLACK}$ = 383/586 = .6536. The comparison is then between $\hat{p}_{cg}$ and the group's representation rate, $\hat{p}_g$, in the district. Thus, $\hat{p}_{EMR,\ BLACK}$ is to be compared with $\hat{p}_{BLACK}$ = 0.4989. The appropriate standard error is $(\hat{p}_g(1-\hat{p}_g)/\text{category total})^{1/2}$ (in the example SE = $(0.4989 \times 0.5011/586)^{1/2}$ = .0207). Then $z_{cg} = (\hat{p}_{cg} - \hat{p}_g)/SE$ (in the example, $z_{EMR,\ BLACK}$ = (.6536 - 0.4989/.0207 = 7.49). As before we would specify a critical value CRIT and declare that disparity (direction not specified) exists for a particular group and category if the z_{cg}-score is, in magnitude greater than CRIT, that is, if z_{cg} > CRIT or z_{cg} < -CRIT (the former indicates overrepresentation, the latter underrepresentation). Under the assumption that the expected proportion of group g in each category is p_g we choose the number CRIT to control the probability, α, of falsely declaring some disparity (type I error in traditional statistical parlance). For example, for α equal to 0.025, CRIT is equal to 3.22 and for α equal to 0.01, CRIT is equal to 3.45; both represent slightly conservative choices.

If we apply the above routine to Table 3 with α = .10 and CRIT = 2.81 then white children are overrepresented in categories ESC, ED, NI, PI, SPCLASS, and underrepresented in EMR; black children are overrepresented in EMR, and underrepresented in ESC, PI, SPCLASS; and hispanic children are

TABLE 3

DISTRICT CODE: 1

CATEGORY	WHITES COUNT	WHITES Z-SCORE	BLACKS COUNT	BLACKS Z-SCORE	HISPANICS COUNT	HISPANICS Z-SCORE
ESC	525	17.29	475	-8.50	256	-6.37
EMR	88	-3.92	383	7.53	115	-4.76
PI	42	3.60	36	-4.27	41	1.44
ED	110	4.81	177	0.88	52	-5.37
NI	206	7.10	298	-0.67	111	-5.74
SPCLASS	5	7.78	7	-6.97	6	0.63
TMR	36	1.11	76	7.53	115	-4.76
TOTAL POPULATION (PROPORTION)	7,046	(0.22)	16,181	(0.50)	9,257	(0.29)

TABLE 4

CODING OF CATEGORIES

Category	Code
Educably Mentally Retarded	EMR
Trainable Mentally Retarded	TMR
Eligible for Speech Correction	ESC
Emotionally Disturbed	ED
Neurologically Impaired	NI
Perceptually Impaired	PI
Special Class	SPCLASS

overrepresented in none and underrepresented in EMR, ESC, ED, and NI. Application of the 80% rule would not alter the results. Although it does not occur in this data set, it is possible that a group may be found overrepresented in a category but neither of the other two be found underrepresented, a result of a strong cummulative underrepresentation of the other two groups but not of each separately. A study of the results of the above analysis reveals patterns that may be of interest. One can only speculate, for example, about why hispanics are underrepresented in a number of categories and overrepresented in none, why blacks are generally underrepresented in categories where whites are overrepresented, or why blacks are overrepresented while whites do not appear underrepresented in EMR though hispanics are underrepresented.

Simultaneous Comparisons: An Example of Screening Several Districts over Several Categories and Groups

Now let us extend our methods to screening several districts in which the potential disparity over several categories and groups exists. We will use this methodology, comprising our final recommendation, on a test data set thus portraying a routine screening episode. For this demonstrative purpose, the data base consists of ten local educational districts, including the one previously analyzed. These ten districts are the largest in child population from all school districts of one governmental state and thus raise the likelihood that all categories and groups are represented in each district. These nine additional districts are presented in Table 5.

The proposed procedure enabling us to screen several districts while

TABLE 5

DATA FOR NINE DISTRICTS

DISTRICT CODE: 2

CATEGORY	WHITES COUNT	WHITES Z-SCORE	BLACKS COUNT	BLACKS Z-SCORE	HISPANICS COUNT	HISPANICS Z-SCORE
ESC	451	6.86	2,081	-23.35	786	4.22
EMR	57	-4.83	836	1.33	141	5.62
PI	26	4.92	51	-8.03	27	-3.85
ED	32	-5.55	658	3.20	95	-5.96
NI	26	0.22	196	-0.18	27	-3.85
SPCLASS	2	-4.85	241	6.71	7	-6.99
TMR	20	-1.32	186	-3.49	58	0.50
TOTAL POPULATION (PROPORTION)	5,574	(0.10)	44,051	(0.79)	11,528	(0.21)

DISTRICT CODE: 3

CATEGORY	WHITES COUNT	WHITES Z-SCORE	BLACKS COUNT	BLACKS Z-SCORE	HISPANICS COUNT	HISPANICS Z-SCORE
ESC	219	4.44	817	3.14	55	-8.68
EMR	17	-1.11	107	1.12	19	0.32
PI	31	0.22	155	2.62	10	-3.65
ED	79	2.34	296	0.97	32	-3.67
NI	80	4.86	214	-1.44	26	-3.12
SPCLASS	30	3.51	72	-1.71	8	-2.09
TMR	18	3.52	32	-2.20	6	-0.75
TOTAL POPULATION (PROPORTION)	2,325	(0.15)	10,763	(0.71)	2,196	(0.14)

DISTRICT CODE: 4

CATEGORY	WHITES COUNT	WHITES Z-SCORE	BLACKS COUNT	BLACKS Z-SCORE	HISPANICS COUNT	HISPANICS Z-SCORE
ESC	19	11.88	431	-8.29	10	0.85
EMR	3	2.95	138	-2.56	3	1.14
PI	3	4.21	85	-0.57	0	-0.98
ED	6	4.63	242	-2.63	3	0.16
NI	3	3.36	117	-1.58	1	-0.28
SPCLASS	0	-0.67	99	1.25	0	-1.05
TMR	2	4.33	37	-3.06	1	0.85
TOTAL POPULATION (PROPORTION)	58	(0.00)	16,838	(0.98)	143	(0.01)

TABLE 5 (Continued)

DATA FOR NINE DISTRICTS

DISTRICT CODE: 5

CATEGORY	WHITES COUNT	WHITES Z-SCORE	BLACKS COUNT	BLACKS Z-SCORE	HISPANICS COUNT	HISPANICS Z-SCORE
ESC	84	3.67	366	-6.41	236	4.62
EMR	51	3.70	207	-3.86	115	1.84
PI	97	9.42	265	-4.39	116	-1.16
ED	36	7.01	95	-0.21	17	-4.16
NI	32	6.14	77	-2.59	32	-1.05
SPCLASS	NR	NR	NR	NR	NR	NR
TMR	20	4.86	50	-1.61	18	-1.31
TOTAL POPULATION (PROPORTION)	1,601	(0.08)	12,439	(0.65)	5,092	(0.27)

DISTRICT CODE: 6

CATEGORY	WHITES COUNT	WHITES Z-SCORE	BLACKS COUNT	BLACKS Z-SCORE	HISPANICS COUNT	HISPANICS Z-SCORE
ESC	299	4.05	187	-7.14	309	3.27
EMR	26	-5.36	115	6.73	55	-1.58
PI	40	-1.77	82	3.91	41	-2.23
ED	65	3.08	65	1.97	23	-4.82
NI	37	0.67	38	-0.17	34	-0.49
SPCLASS	159	0.59	186	0.93	149	-1.52
TMR	14	-1.50	28	1.46	21	-0.01
TOTAL POPULATION (PROPORTION)	4,413	(0.31)	5,080	(0.36)	4,759	(0.34)

DISTRICT CODE: 7

CATEGORY	WHITES COUNT	WHITES Z-SCORE	BLACKS COUNT	BLACKS Z-SCORE	HISPANICS COUNT	HISPANICS Z-SCORE
ESC	660	-1.13	9	-0.05	12	1.75
EMR	44	-0.11	1	0.72	0	-0.63
PI	236	-6.01	12	4.62	7	2.57
ED	114	-1.84	2	0.31	4	2.41
NI	103	-0.89	3	1.31	1	-0.14
SPCLASS	478	-0.05	7	0.16	5	-0.11
TMR	36	-0.11	1	0.72	0	-0.63
TOTAL POPULATION (PROPORTION)	16,753	(0.96)	231	(0.01)	184	(0.01)

NR = NONE REPORTED.

TABLE 5 (Continued)

DATA FOR NINE DISTRICTS

DISTRICT CODE: 8

CATEGORY	WHITES COUNT	WHITES Z-SCORE	BLACKS COUNT	BLACKS Z-SCORE	HISPANICS COUNT	HISPANICS Z-SCORE
ESC	340	-3.36	11	0.32	13	0.04
EMR	37	-0.29	4	2.74	0	1.24
PI	327	-3.83	21	3.69	4	1.14
ED	162	1.12	1	-1.68	2	0.90
NI	97	1.30	1	-1.05	0	-0.80
SPCLASS	1,055	0.02	30	-0.01	7	0.04
TMR	46	-0.29	1	-0.28	1	1.24
TOTAL POPULATION (PROPORTION)	12,639	(0.97)	360	(0.03)	85	(0.01)

DISTRICT CODE: 9

CATEGORY	WHITES COUNT	WHITES Z-SCORE	BLACKS COUNT	BLACKS Z-SCORE	HISPANICS COUNT	HISPANICS Z-SCORE
ESC	NR	NR	NR	NR	NR	NR
EMR	60	-2.58	321	2.93	171	-1.16
PI	NR	NR	NR	NR	NR	NR
ED	8	0.78	27	1.60	7	-2.29
NI	28	1.96	85	2.57	22	-4.19
SPCLASS	NR	NR	NR	NR	NR	NR
TMR	12	1.62	27	-0.14	14	-1.06
TOTAL POPULATION (PROPORTION)	3,724	(0.15)	13,098	(0.52)	8,401	(0.33)

DISTRICT CODE: 10

CATEGORY	WHITES COUNT	WHITES Z-SCORE	BLACKS COUNT	BLACKS Z-SCORE	HISPANICS COUNT	HISPANICS Z-SCORE
ESC	NR	NR	NR	NR	NR	NR
EMR	58	1.12	1	-0.60	0	-1.03
PI	531	-0.58	12	-1.24	18	2.56
ED	178	-0.64	10	1.80	1	-1.30
NI	145	0.12	4	-0.29	3	0.18
SPCLASS	384	3.72	3	-2.59	0	-2.65
TMR	44	0.81	1	-0.32	0	-0.90
TOTAL POPULATION (PROPORTION)	13,467	(0.95)	430	(0.03)	252	(0.02)

NR = NONE REPORTED.

detecting patterns of disparity within districts breaks into two stages. The first stage focuses on which, if any, districts have one or more disparate representation rates, those that do are flagged. Following this, the second stage focuses on the pattern of disparities only within those districts having been detected in the first stage. In our example, the first stage amounts to performing 210 simultaneous statistical tests. Using a ten percent type I error of falsely declaring some representation rate (and hence the district it belongs to) disparate we find that CRIT = 3.45. While the second stage amounts to performing 21 simultaneous statistical tests. Once again, using a ten percent type I error we find that CRIT = 2.81.

The first stage screening of districts results in disparities occurring in all ten districts; the degree of disparity in a district, according to the number of incidents, is not evenly distributed over all districts. The number of incidents is distributed as follows: two districts with one disparity each, one district with three, one with four, three with six each, one with nine, one with 14, and finally one with 15. In summary, the majority of districts have between four and nine disparities, with the least and most being one and 15 disparities respectively. Consequently, although these ten districts are similar with regard to size, they are quite dissimilar with regard to the number of disparities in each district.

The second stage screening of categories within each district yields the summary portrayed in Figure 1. Letting □ denote overrepresentation, ▭ denote underrepresentation and - denote neither, we obtain a visual portrayal of the qualitative nature of disparity in a district with regard to categories and groups. The ordering of the symbols for each category and district correspond respectively to white, black, and hispanic. In addition, Figure 1 includes

group representation rates in the total child populations of each district, enabling us to conjecture a possible cause of the observed patterns of disparities. Returning to this point later, we discuss the degree of disparity over categories and then the degree, as well as the direction, of the disparaties over groups.

- - - - - - - - - - -

Figure 1 about here

- - - - - - - - - - -

The ESC category has the largest number of incidents of disparity, 17. The remaining categories, in descending order, have 14, 13, 9, 8, 8, and 6 for EMR, PI, ED, NI, SPCLASS and TMR, respectively. Thus with regard to incidents of disparity, ESC, EMR and PI appear to be the more active categories, while the remaining categories, although not as abused, still play a significant role in disparate placement.

Counting the number of districts with one or more disparities by category enables us to obtain a different measure of the degree of disparity. The number of districts, in descending order are 8, 8, 7, 6, 5, 5, and 4 for EMR, PI, ESC, NI, ED, TMR, and SPCLASS, respectively. Thus disparity in one or more categories is not concentrated in primarily two or three districts, but appears to be spread out over a large majority of the ten districts.

Counting the joint occurrences of disparity in multiple categories enables us to determine the association; if any, between disparate placement in categories. Although we will not exhaust all possible contingencies, we will note the most frequent ones. Simultaneous disparity in PI and ESC placements occurs in seven districts, PI and EMR in six, and EMR and ESC in five. In

FIGURE 1: SUMMARY OF SIMULTANEOUS SIGNIFICANCE TESTING ON TEN DISTRICTS

addition, simultaneous disparity in PI, EMR and ESC placements occurs in five districts. An underlying process common to all the districts could possibly cause this observed association between PI, EMR, and ESC placements; more detailed data than we have would be needed to investigate this topic further.

We now summarize disparities over the groups, white children being discussed first followed by black and then hispanic children. In addition, we will also distinguish between disparities with regard to overrepresentation and underrepresentation.

White children are overrepresented by descending order in ESC, PI, NI, EMR, TMR, ED, and SPCLASS in 6, 4, 4, 3, 3, 3, and 3 districts, respectively. Simultaneous overrepresentation in ESC and PI occurs in four districts while simultaneous overrepresentation in ESC, PI, and EMR occurs in two districts. In addition, white children are underrepresented in EMR, PI, ED,SPCLASS, TMR, ESC, and NI in 4, 2, 1, 1, 0, 0, and 0 districts, respectively. Simultaneous underrepresentation in EMR and PI occurs in two districts.

Black children are overrepresented by descending order in EMR, PI, ESC, ED, SPCLASS, TMR, NI, in 4, 3, 1, 1, 1, 0, and 0 districts, respectively. Simultaneous overrepresentation in EMR and PI occurs in two districts. In addition, black children are underrepresented in ESC, PI, TMR, EMR, SPCLASS, ED, and NI in 5, 3, 2, 1, 1, 0, and 0 districts, respectively. Simultaneous underrepresentation in ESC and PI occurs in three districts.

Hispanic children are overrepresented in ESC in three districts but not in any other category. In addition, hispanic children are underrepresented by descending order in ED, NI, EMR, ESC, SPCLASS, TMR, and PI in 4, 4, 2, 2, 1, and 1 districts, respectively. Simultaneous underrepresentation in ED and NI

occurs in three districts. All the above disparities are graphically compared in Figure 2.

In summary, white children are overrepresented in 26 incidents, black children in 10, and hispanic children in 3, while white children are underrepresented in eight incidents, black children in 12, and hispanic children in 16.

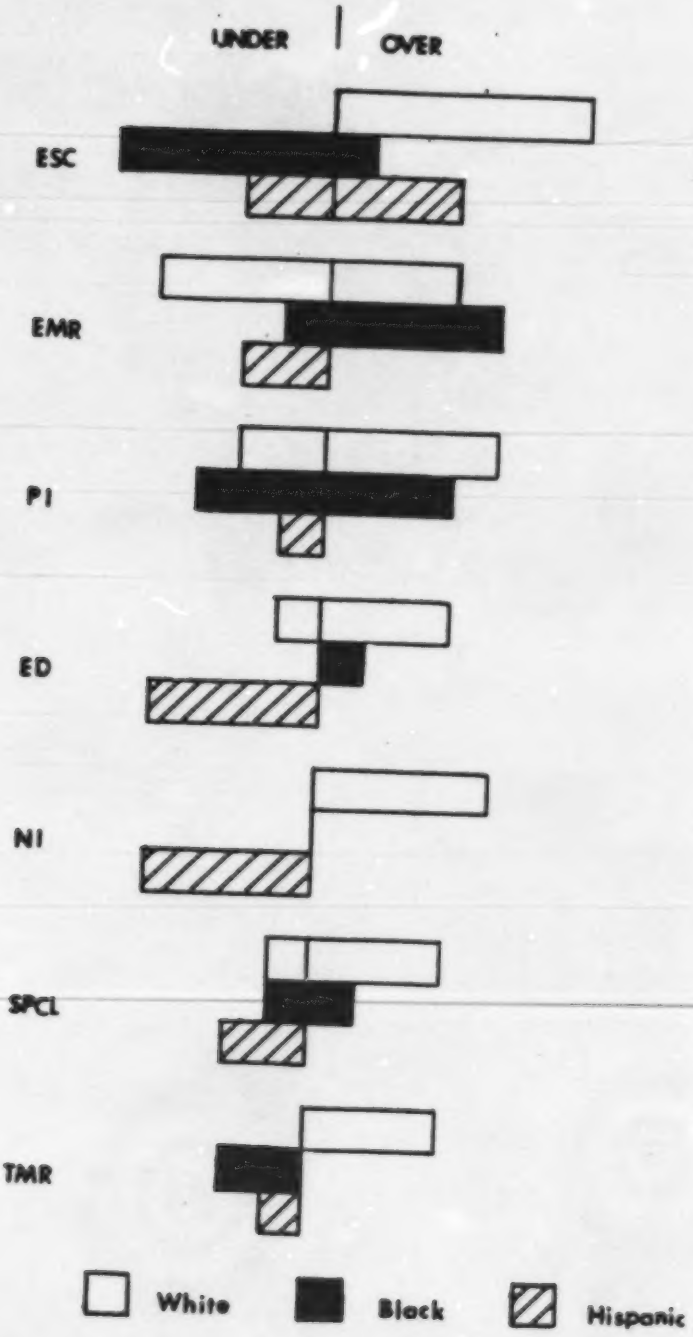
- - - - - - - - - - -

Figure 2 about here

- - - - - - - - - - -

These results raise several questions, the primary ones being why are ESC, EMR, and PI placements more abused than the others and why are white children overrepresented in ESC, PI, and NI and underrepresented in EMR, while black children are overrepresented in EMR and PI and underrepresented in ESC? In addition, why do white children tend to be overrepresented over all districts while hispanic children tend to be underrepresented over all districts? These questions cannot be properly answered with the available data; however, some observations which are related to the representation of the groups in the districts can be made. For example, in the six districts in which white children are overrepresented in ESC, they also comprise the minority group (the percentages are approximately 0.4, 8.4, 10.0, 15.2, 21.7, and 31.0); whereas, in the one district in which they are underrepresented in ESC, they comprise the majority group (at 96.6 percent). Other conjectured relationships such as this could be drawn, however, it is not our purpose in this article to explore them but rather to demonstrate our methodology. The purpose of the digression, however, does serve to indicate the value of our

FIGURE 2: SUMMARY OF SCREENING TEN DISTRICTS*



* One squares is equivalent to one incident of disparity.

approach in that our methodology does reveal systematic disparaties that tend to fit preconceived notions of the nature and direction of disparate placements while at the same time revealing unsuspected patterns of disparity.

Appendix

We may use an analogy in order to find a critical value that several representation rates must exceed in magnitude in order to be declared significant while maintaining control over the type I error. Suppose we have one coin whose probability of heads we may adjust so that upon flipping the coin 100 times, for example, the probability of at least one head is ten percent. We draw the correspondence between comparing one representation rate to the critical value and flipping the coin once, and between the representation rate exceeding in magnitude the critical value and the occurrence of a head on the flip. Needless to say, the probability of heads must be very small, and correspondingly the magnitude of the critical value large, in order to have only a ten percent chance of getting at least one head in 100 tosses. In fact, this value can be readily determined to be approximately 0.00105305031. Fortunately there is a formula for this probability, p: $p=1-\sqrt[M]{1-\alpha}$, where α is the probability of at least one head in M tosses of the coin. For M equal to 21, p is equal to 0.005 and for M equal to 210, p is equal to 0.0005. The value of p determines the critical value, CRIT. Critical values for screening for both over- and underrepresentation that yield these two values of p are CRIT=2.81 and CRIT=3.45. These values are obtainable from a table of the percentiles of a standard normal probability distribution; the respective percentiles corresponding to p equal to 0.005 and

0.0005 are 99.75 and 99.975. A convenient approximation formula for a critical value corresponding to values of p less than or equal to 0.005 is $CRIT = [(1-p/2)^{0.135} - (p/2)^{0.135}] / 0.198$.

Conclusion

The uniform use of statistical methods to detect disproportionate placement in special education programs can lead to serious errors in district and state-level compliance monitoring. These errors can result from a variety of factors. These factors include using statistical tests when the assumptions for random sampling are grossly violated, using the 80% standard alone, comparing enrollment rates instead of representation rates, and computing statistical tests separately over many school districts without accounting for the likelihood of finding significant results due to repeated comparisons. Given these pitfalls, we have offered a method that can be effectively used to identify the existence of disparities in different diagnostic categories across many school districts at once.

REFERENCES

Ashurst, D. and Meyer, C. E. Social system and clinical model in school identification of the educable mentally retarded. In R.K. Eyman, C.E. Meyer, and G. Tarjan (eds.), Social Behavioral Studies in Mental Retardation. Washington, D.C., American Association on Mental Deficiency, 1973.

Bickel, P. J., Hammel, E. A., and O'Connell, J. W. Sex bias in graduate admissions: data from Berkeley. Science, 137, 1975, 398-404.

Connecticut v. Teal. U.S. Law Weekly, 50, 1982, 4716 (U.S. June 21, 1982).

Hazelwood School District v. United States, 433 U.S. 99, 1977.

Heller, K., Holtzman, W., and Messick, S. Placing children in special education: Strategy for equity. Washington, D.C.: National Academy Press, 1982.

International Brotherhood of Teamsters v. United States, 431, U.S.324, 1977.

Kruskal, W. Communicated in Statistics and Public Policy. Fairley, A. and Mosteller, F., 1977, 127-130.

Lambert, N. Psychological evidence in Larry P. v. Wilson Riles: An evaluation by a witness for the defense. American Psychologist, 36, 1981, 937-952.

Larry P. et al., v. Wilson Riles et al. United States District Court, Case No. C- -2270RFP, 1974, 1979.

Lora v. Board of Education of the City of New York. Opinion, U.S. District Court for the Eastern District of New York (75 Civ. 917), July 1, 1979.

Mattie T. et al. v. C. E. Halladay et al. C.A. No. DC-75-31-S United States District Court for the Northern District of Mississippi, 1979.

REFERENCES

Meier, P., Sacks, J., and Zabell, S. What happened in Hazelwood, statistics, employment discrimination, and the 80% rule. To appear in the American Bar Foundation Journal.

Miller, R. G. Simultaneous statistical inference, Second edition. New York: Springer-Verlag, 1981.

Parents in Action on Special Education (PASE) v. Hannon. Opinion, U.S. District Court for the Northern District of Illinois (No. 74-C-3586), July 11, 1980.

Prasse, D. The Larry P. decision. Special issue of School Psychology Review, 9, (2), 1980.

Reschly, D. J. SOMPA: A symposium. Special issue of School Psychology Review, 8, (2), 1979.

Uniform Guidelines on Employment Selection Procedures, 1978, 43FR38290-38318, (August 25, 1978).

U.S. Department of Education. Assistance to the states for education of the handicapped. Federal Register, 42, 1977, 65082-65085.

U.S. Office of Education. Education of handicapped children: Implementation of Part B of the Education of the Handicapped Act. Federal Register, 42, 1977, 42474-42518.